Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **AUG 2004**	2. REPORT TYPE	3. DATES COVERED **00-00-2004 to 00-00-2004**

4. TITLE AND SUBTITLE **The Semantics of Ontology Alignment**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Lockheed Martin Advanced Technology Laboratories,3 Executive Campus,Cherry Hill,NJ,08002**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Proceedings of the 2004 Performance Metrics for Intelligent Systems Workshop (PerMIS ?04), Gaithersburg, MD on August 24-26 2004

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**3**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

The Semantics of Ontology Alignment

Todd C. Hughes and Benjamin C. Ashpole
Lockheed Martin Advanced Technology Laboratories
Cherry Hill, NJ
{thughes, bashpole}@atl.lmco.com

ABSTRACT

Ontology alignment is a foundational problem area for semantic interoperability. We discuss the complexity faced by automated alignment solutions and describe an ontology-based approach for describing and evaluating alignments.

KEYWORDS: ontology, ontology alignment, ontology mapping, Semantic Web

1 THE CHALLENGE OF ONTOLOGY ALIGNMENT

The vision of semantic interoperability, the fluid sharing of services and digitalized knowledge, is often thought to hinge on a common, formal language that machines can somehow understand. However, protocols and data formats such as XML tags and schemas have proven to be inadequate solutions primarily because the burden of meaning is still on humans, who still must learn implicit semantics of foreign systems in order to make them work with their native systems. Semantic Web languages like RDF and OWL begin to ameliorate the problem by adding explicit semantic relationships and logical constraints between elements (i.e., classes, properties, and restrictions) in the form of ontologies, an extension of schemas. However, programs that read OWL documents that conform to a particular ontology cannot understand other OWL documents conform to a different ontology *unless* there is an explicit mapping between the ontologies. Creating this mapping is the alignment problem, and solving it is the first step to semantic interoperability.

Alignment between ontologies is a critical challenge for semantic interoperability. There are $(n*m)$ possible *individual*, undirected alignments for ontology graphs of size *n* and *m*. Optimal graph matching algorithms run in exponential time due to the NP-complete nature of the search space. For large ontologies with tens of thousands of elements, purely manual alignment methods are clearly impractical [1], and semi-automated approaches are not suitable for real-time applications.

Semantic interoperability requires fully automated ontology alignment approximation techniques. This cannot be accomplished solely by lexical comparison between element names in different ontologies, since names (like tags) can be abbreviations, acronyms, phrases, in different languages, misspelled, or used in unexpected, jargon-specific ways. In addition, the size, structure, and scope of ontologies must be considered. There is no guarantee that two ontologies in the same domain will have terms that all *precisely* and *completely* overlap: in one ontology, an element name might be equivalent to several—or none—in another. Clearly, alignment techniques must be sensitive to a number of ontology features to find corresponding elements. [4]

A number of prototype ontology alignment applications have been developed to meet this challenge. However, it is difficult to assess the efficacy of these tools because their developers each use their own alignment formats, test data sets, and evaluation metrics. Do et al [2] have made a notable effort to compare alignment tools using standard metrics, but at this point in alignment research there is still no formal, broadly used language to describe the output of an aligner and to judge the value of one aligner relative to another.

2 ALIGNMENT SEMANTICS

We have developed, appropriately enough, a set of ontologies intended to capture the semantics for relevant metrics for automated ontology operations, including ontology alignment. These ontologies are part of an ongoing effort to focus the ontology alignment community on canonical set of challenge problems, research objectives, and evaluation criteria. Here we describe some of the classes and properties of our ontologies, which are available on our website [1].

2.1 Alignment and Equivalence

Alignment is distinct from equivalence for at least two fundamental reasons. First, an ontology alignment provides only a relation between ontology elements: any particular element alignment will depend on the alignments between other elements. An ontology alignment is the most stable set of element alignments, at least in the opinion of the aligner. This leads to a second difference, namely, that element alignments can (and often do) have degrees of confidence associated with them. That is, the aligner cannot say with certainty any particular alignment is true, only that it is the most probable alignment given other alignments.

The differences suggest that current Semantic Web terms for expressing equivalence, such as `owl:sameClassAs`, `owl:samePropertyAs`, and `owl:sameAs` are not adequate for expressing alignments.

These properties are intended to capture logical, not relative, equivalence. At this time, there are no broadly accepted semantics for describing the uncertainty of equivalence statements made using these properties, nor is it clear there should be.

2.2 Alignment File

These considerations have led to a different approach to describing alignments formally. We describe an ontology alignment in a Semantic Web document called an `AlignmentFile`. An Alignment File declares instances of the class `Alignment`, where each instance states that an element from one ontology (`elementA`) corresponds to an element from the other ontology (`elementB`) with some probability (`alignmentConfidence`). An example alignment is shown in Figure 1.

```
:Alignment1 a ao:Alignment;
      ao:elementA
      <someOntologyA#ClassA>;
      ao:elementB
      <someOntologyB#ClassB>;
      ao:alignmentConfidence "0.5".
```

Figure 1. An example instance of an Alignment, shown in the N3 language with simplified URI's. [1] The classes and properties are all defined in the Alignment Ontology (referred via the ao: prefix).

The Alignment File format easily allows for 1-n and n-1 element alignments. Should they become useful in the future, it also allows for more unusual alignments, such alignments between a class and a property or a (group of) instances and a class.

2.3 Alignment Evaluation File

Alignment Files deliver the output of ontology alignment algorithms. To assess the performance of that algorithm, one may compare the output to a document that contains the correct (or best) element alignments for the ontologies in question. We refer to this document using the property `trueAlignment`, and we create this document by hand. When an automated grader compares an alignment file to a true alignment file, it delivers another document called an `AlignmentEvaluationFile`.

There are two broad categories of metrics to consider when evaluating an alignment: experiment metrics and performance metrics. The first category concerns the behavior of the aligner in the experiment, independent of the true ontology alignment. Experiment metrics include but are not limited to:

- `meanGlobalCardinality`: For 1:n alignments between elements between Ontology A and Ontology B, this property expresses the average value for n. (Based on Do et al [2] local/global cardinality metric.)
- `sdGlobalCardinality`: For 1:n alignments between elements between Ontology A and Ontology B, this property expresses the standard deviation value for n. (Also based on Do et al [2] local/global cardinality metric.)
- `unalignedElements`: The number of elements in Ontology A for which no corresponding element in Ontology B has been found.
- `alignedProportion`: The proportion of elements from Ontology B that were aligned to elements from Ontology A.
- `uniqueElements`: The proportion of resources not shared (i.e, having different URIs) between Ontology A and Ontology B.
- `alignmentChallenge`: The proportion of unique elements between Ontology A and Ontology B to the total number of elements in Ontology A and Ontology B.

The second category of metrics concerns the correctness of the element alignments contained in the alignment file. A number of these metrics are derivative of well-known metrics from the information retrieval domain.

- `truePositives`: The number of correct alignments an alignment file contains.
- `falsePositives`: The number of incorrect a alignments an alignment file contains.
- `falseNegatives`: The number of correct alignments missed in an alignment file.
- `precision`: The proportion of correct alignments among those found, (truePositives / (truePositives + falsePositives).
- `recall`: The proportion of correct alignments found (truePositives / (truePositives + falseNegatives)).
- `fMeasure`: The harmonic mean of precision and recall (2*(precision*recall)/(precision + recall)).
- `alignmentPerformance`: Indicates performance given the proportion of overlapping resources between Ontology A and Ontology B (alignmentChallenge * fMeasure).

All of the above performance metrics with the exception of last one are borrowed from Do et al [4]. A partial example of an Alignment Evaluation File is shown in Figure 2. The metrics provide a fairly comprehensive account of the performance of an alignment algorithm, while the ontological framework allows the addition of new metrics as needed.

```
<AlignmentEvaluation123.n3>
      a ae:AlignmentEvaluationFile;
      oe:evaluates
<AlignmentFile123.n3>;
      ae:trueAlignment
<TrueAlignmentAB.n3>;
      oe:grader <Grader1.n3>;
      ae:meanGlobalCardinality "0.5";
      ae:sdGlobalCardinality "0.5";
      ae:unalignedElements "0.5";
      ae:alignmentProportion "0.5";
      ae:uniqueElements "0.5";
      ae:alignmentChallenge "0.8";
      ae:truePositives "0.8";
      ae:falseNegatives "0.0";
      ae:precision "0.8";
      ae:recall "0.8";
      ae:fMeasure "0.8".
      ae:alignmentPerformance "0.64".
```

Figure 2. Partial Example Alignment Evaluation File with Simplified URIs. Prefixes refer to the Alignment Evaluation Ontology (ae:) and an "upper" Ontology Operation Evaluation Ontology (oe:).

3 CONCLUSION

The purpose of creating these ontologies is not only to facilitate our own experimentation with alignment algorithms, but also to facilitate greater collaboration among members of the ontology alignment research community. With a common representational scheme for stating and evaluating alignments, it becomes significantly easier to compare alignment algorithms. Of course, in addition to this framework canonical data sets are also needed to ensure fair and accurate comparisons.

To that end, we have made these ontologies freely available on our website, which also includes sample data sets and an Experiment Set Platform for administering ontology alignment experiments. We have collaborated with NIST to establish an ontology alignment competition based on the model of the Text Retrieval Conference (TREC), called the Information Interpretation and Integration Conference (I^3CON). This event will be the first systematic comparison of ontology alignment algorithms.

Finally, it bears mentioning that ontology alignment is not valuable for its own sake, but is worthwhile only in the service of some other function that requires it. We envision considerable value in automated ontology alignment capabilities for agents that semantically interoperate with heterogeneous (particularly legacy) data systems. As such, the ontologies for ontology alignment should grow to encompass semantic interoperation use cases. These new concepts should allow us to articulate in a formal way the impact of ontology alignment on agent mission success.

4 REFERENCES

[1] Berners-Lee, Timothy. "Getting into the Semantic Web and RDF using N3." http://www.w3.org/2000/10/swap/Primer.html

[2] Do, Hong-Hai, Sergey Melnik, and Erhard Rahm. "Comparison of Schema Matching Evaluations." *Proceedings of the 2nd Int. Workshop on Web Databases*. German Informatics Society, 2002. Available at: http://citeseer.ist.psu.edu/do02comparison.html

[3] McGuinness, Deborah L., Richard Fikes, James Rice, and Steve Wilder. "An Environment for Merging and Testing Large Ontologies." *Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR2000)*. Breckenridge, Colorado, USA. April 12-15, 2000. Available at: http://www.ksl.stanford.edu/software/chimaera/

[4] Rahm, Erhard, and Phillip A. Bernstein. "On Matching Schemas Automatically." Technical Report for Univ. Leipzig, Dept. of Computer Science, Database Group. 2001. Available at: http://dol.uni-leipzig.de/pub/2001-5

[5] The Ontology Alignment Source, http://www.atl.lmco.com/projects/ontology/

[6] Information Interpretation and Integration Conference, http://www.atl.lmco.com/projects/ontology/i3con.html